

What is claimed is:

1 1. A speech recognition system for transforming an acoustic signal into a stream of phonetic
2 estimates, comprising:

3 a frequency analyzer for receiving the acoustic signal and producing as an output a short-
4 time frequency representation of the acoustic signal;

5 a novelty processor for receiving the short-time frequency representation of the acoustic
6 signal, separating one or more background components of the representation from one or more
7 region-of-interest components of the representation, and producing a novelty output including the
8 region of interest components of the representation according to one or more novelty parameters;

9 an attention processor for receiving the novelty output and producing a gating signal as a
10 predetermined function of the novelty output according to one or more attention parameters;

11 a coincidence processor for receiving the novelty output and the gating signal, and
12 producing a coincidence output that includes co-occurrences between samples of the novelty
13 output over time and frequency, wherein the coincidence output is selectively gated as a
14 predetermined function of the gating signal, so as to produce a gated coincidence output
15 according to one or more coincidence parameters; and,

16 a vector pattern recognizer and a probability processor for receiving the gated coincidence
17 output and producing a phonetic estimate stream representative of acoustic signal.

1 2. A speech recognition system according to claim 1, wherein the short-time frequency
2 representation of the audio signal includes a series of consecutive time instances, each
3 consecutive pair separated by a sampling interval, and each of the time instances further includes
4 a series of discrete Fourier transform (DFT) points, such that the short-time frequency
5 representation of the audio signal includes a series of DFT points.

1 3. A speech recognition system according to claim 2, wherein for each DFT point, the
2 novelty processor (i) calculates a first average value across a first predetermined frequency range
3 and a first predetermined time span, (ii) calculates a second average value across a second
4 predetermined frequency range and a second predetermined time span, and (iii) subtracts the
5 second average value from the first average value so as to produce the novelty output point.

1 4. A speech recognition system according to claim 3, wherein the first frequency range, the
2 first time span, the second frequency range and the second time span are each a function of one
3 or more of the novelty parameters.

1 5. A speech recognition system according to claim 3, wherein the first predetermined
2 frequency range is substantially centered about a frequency corresponding to DFT point, and the
3 first predetermined time span is substantially centered about an instant in time corresponding to
4 the DFT point.

1 6. A speech recognition system according to claim 3, wherein the first predetermined
2 frequency range is substantially smaller than the second predetermined frequency range.

1 7. A speech recognition system according to claim 3, wherein the first predetermined time
2 span is substantially smaller than the second predetermined time span.

1 8. A speech recognition system according to claim 3, wherein the second predetermined
2 time span is large relative to the second predetermined frequency range.

1 9. A speech recognition system according to claim 3, wherein the second predetermined
2 frequency range is large relative to the second predetermined time span.

1 10. A speech recognition system according to claim 3, wherein for each DFT point, the
2 novelty processor further calculates one or more additional novelty outputs, and each additional
3 novelty output is defined by characteristics including a distinct first frequency range, first time
4 span, second frequency range and second time span, each characteristic being a function of one
5 or more of the novelty parameters.

1 11. A speech recognition system according to claim 2, wherein the coincidence output
2 includes a sum of products of novelty output points over two sets of novelty output points.

1 12. A speech recognition system according to claim 11, wherein the two sets of DFT points
2 includes a first set of novelty output points corresponding to a first instant in time and a second
3 set of novelty output points corresponding to a second time instance.

1 13. A speech recognition system according to claim 11, wherein the two sets of novelty
2 output points all correspond to a single time instance.

1 14. A speech recognition system according to claim 11, wherein the coincidence processor
2 performs the sum of products of novelty output points over two sets of novelty output points
3 according to one or more selectably variable coincidence parameters including time duration,
4 frequency extent, base time, base frequency, delta time, delta frequency, and combinations
5 thereof.

1 15. A speech recognition system according to claim 2, wherein each of the time instances
2 further includes an energy value in addition to the series of DFT points.

1 16. A speech recognition system according to claim 15, wherein the attention processor (i)
2 compares the energy value to a predetermined threshold value according to a comparison
3 criterion, so as to produce an energy threshold determination, and (ii) produces the gating signal
4 as a predetermined function of the threshold determination.

1 17. A speech recognition system according to claim 16, wherein the one or more attention
2 parameters include the predetermined threshold value, the comparison criterion and the
3 predetermined function of the threshold determination.

1 18. A speech recognition system according to claim 1, wherein the novelty parameters, the
2 attention parameters and the coincidence parameters are selected via a genetic algorithm.

1 19. A speech recognition system for transforming a short-time frequency representation of an
2 acoustic signal into a stream of coincidence vectors, comprising:

3 a novelty processor for receiving the short-time frequency representation of the audio
4 signal, separating one or more background components of the signal from one or more region of
5 interest components of the signal, and producing a novelty output including the region of interest
6 components of the signal according to one or more novelty parameters;

7 a coincidence processor for receiving the novelty output and the gating signal, and
8 producing a coincidence vector that includes data describing co-occurrences between samples of
9 the novelty output over time and frequency according to one or more coincidence parameters.

1 20. A speech recognition system according to claim 19, further including an attention
2 processor for receiving the novelty output and producing a gating signal as a predetermined
3 function of the novelty output according to one or more attention parameters, wherein the
4 coincidence output is selectively gated as a predetermined function of the gating signal, so as to
5 produce a gated coincidence output according to one or more coincidence parameters.

1 21 A speech recognition system according to claim 19, wherein the novelty parameters and
2 the coincidence parameters are selected via a genetic algorithm.

1 22. A method of transforming an acoustic signal into a stream of phonetic estimates,
2 comprising:
3 receiving the acoustic signal and producing a short-time frequency representation of the
4 acoustic signal;
5 separating one or more background components of the representation from one or more
6 region of interest components of the representation, and producing a novelty output including the
7 region of interest components of the representation according to one or more novelty parameters;
8 producing a gating signal as a predetermined function of the novelty output according to
9 one or more attention parameters;
10 producing a coincidence output that includes correlations between samples of the novelty
11 output over time and frequency, wherein the coincidence output is selectively gated as a
12 predetermined function of the gating signal, so as to produce a gated coincidence output
13 according to one or more coincidence parameters; and,
14 producing a phonetic estimate stream representative of acoustic signal as a function of the
15 gated coincidence output.

1 23. A method according to claim 22, further including (i) calculating a first average value
2 across a first predetermined frequency range and a first predetermined time span, (ii) calculating
3 a second average value across a second predetermined frequency range and a second
4 predetermined time span, and (iii) subtracting the second average value from the first average
5 value so as to produce the novelty output.

1 24. A method according to claim 22, further including calculating, for each of a plurality of
2 DFT points from the a short-time frequency representation of the acoustic signal, one or more
3 additional novelty outputs, wherein each additional novelty output is defined by characteristics
4 including a distinct first frequency range, first time span, second frequency range and second
5 time span, each characteristic being a function of one or more of the novelty parameters.

1 25. A method according to claim 24, further including performing a sum of products of
2 novelty outputs over two sets of novelty outputs according to one or more selectably variable
3 coincidence parameters including time duration, frequency extent, base time, base frequency,
4 delta time, delta frequency, and combinations thereof.

1 26. A method according to claim 22, further including comparing the energy value to a
2 predetermined threshold value according to a comparison criterion, so as to produce an energy
3 threshold determination, and (ii) producing the gating signal as a predetermined function of the
4 threshold determination.

1 27. A method according to claim 22, further including selecting the novelty parameters, the
2 attention parameters and the coincidence parameters via a genetic algorithm.